AFOSR Grant Final Performance Report

A System Approach to Adaptive Multi-Modal Sensor Designs

Zhigang Zhu (PI)
Department of Computer Science, City College of New York /CUNY
zhu@cs.ccny.cuny.edu

Harvey E. Rhody (Co-PI)

Center for Imaging Science, Rochester Institute of Technology

Email: rhody@cis.rit.edu

Program Managers:
Dr. Douglas Cochran <douglas.cochran@afosr.af.mil>
Dr. Kitt C. Reinhardt <kitt.reinhardt@afosr.af.mil>

Prepared for

Air Force Office of Scientific Research

under Award FA9550-08-1-0199

February 2010

The City College of New York

DEPARTMENT OF COMPUTER SCIENCE
SCHOOL OF ENGINEERING

CONVENT AVE & 138TH ST
NEW YORK, NY 10031

Approved for public release; distribution is unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
February 2010	Final Performance Report	04/01/2008 - 11/30/2009

4. TITLE AND SUBTITLE

A System Approach to Adaptive Multi-Modal Sensor Designs

5a. CONTRACT NUMBER
FA9550-08-1-0199

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Zhigang Zhu, Harvey E. Rhody

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Computer Science Department
City College of New York
Convent Avenue and 138th Street
New York, NY 10031

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr. Douglas Cochran

Air Force Office of Scientific Research

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We propose a system approach to adaptive multimodal sensor designs. This approach is based on the integration of tools for the physics-based simulation of complex scenes and targets, sensor modeling, and multimodal data exploitation. The goal is to reduce development time and system cost while achieving optimal results through an iterative process that incorporates simulation, sensing, processing and evaluation. A Data Process Management Architecture (DPMA) is designed, which is a software system that provides a team development environment and a structured operational platform for systems that require many interrelated and coordinated steps. As a case study, we use an effective peripheral-fovea design as an example. This design is inspired by the biological vision systems for achieving real-time target detection and recognition with a hyperspectral/range fovea and panoramic peripheral view. Under the principle of the system approach and performance-driven sensing, a real multimodal human signature detection sensing design is proposed and a test prototype is developed to capture visual, audio and range information at a large distance.

15. SUBJECT TERMS
Hyperspectral imaging; panoramic vision; system architecture; multimodal sensing; target tracking

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	SAR	30	(212) 650-8799

ACKNOWLEDGEMENTS

In addition to the support of the AFOSR under Award #FA9550-08-1-0199, this work is also supported by in part by AFRL/SN under Award No. FA8650-05-1-1853, ARO under a DURIP Award #W911NF-08-1-0531, and by NSF under grant No. CNS-0551598. The generous supports of AFOSR and the other funding agencies are greatly appreciated.

This is a joint work between the City College of New York (CCNY) and Rochester Institute of Technology (RIT). We would like to thank our team members and collaborators at both institutions during the period of the project: Dr. Yufu Qu, Dr. Jizhong Xiao, Mr. Tao Wang, Mr. Edgardo Molina, Mr. Wai L. Khoo, Mr. Hao Tang, all at CCNY, and Mr. Robert S. Krzaczek at RIT. We would also like to thank Dr. Erik Blasch at the Air Force Research Laboratory (AFRL/RYAA) for his collaboration when our PhD student Mr. Tao Wang and the PI (Dr. Zhigang Zhu) did the summer visit research at AFRL in 2008, and Dr. Ajay Divakaran at Sarnoff for the collaboration in audio-visual event detection.

The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. However, the views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Government.

Contents

1. Introduction 5
1.1. What is needed? 5
1.2. A system approach 6
2. System Architecture 7
3. A Bio-Inspired Sensor Design 9
4. Scene Simulation and Sensor Modeling 12
5. Data Exploitation and Adaptive Sensing 14
5.1. Detection and tracking in peripheral views 15
5.2. Target classification using 3D and HSI fovea 17
5.3. Experimental results 19
6. A Multimodal Sensing Platform with Real Sensors 21
6.1. Problems 21
6.2. Vision-aided surface detection, orientation estimation and laser aiming and focusing 22
6.3. Automatic and adaptive audio-visual signal acquisition and processing 23
7. Conclusion and Discussions 24
References 26
Appendix: Publications under the Support of This Grant 29

1. Introduction

Recently, a great deal of effort has been put into adaptive and tunable multi-spectral or hyper-spectral sensor designs with goals to address the challenging problems of detecting, tracking and identifying targets in highly cluttered, dynamic scenes. Representative large programs include: the DARPA's Adaptive Focal Plane Array (AFPA) Program (Horn 2007), ARL's Advanced Sensor CTA (Goldberg, et al. 2003), and NSF's Center for Mid-Infrared Technologies for Health and the Environment (Kincade 2006). Whereas these efforts mainly focus on semiconductor materials, photonics, and hardware designs, and have created or will soon create novel adaptive multimodal sensors, the sensors that have been designed and are being designed are not up to the expectations for real-world applications. Another piece of novel sensor design will not by itself revolutionarily change this situation. Our work looks into using a system approach to utilize advanced multimodal data exploitation and information sciences for innovative multimodal sensor designs to satisfy the requirements of real-world applications in security, surveillance and inspection.

1.1. What is needed?

Modern forward-looking infrared (FLIR) imaging sensors can achieve high detection and low false alarm rates through the exploitation of the very high spatial resolution available on current generation of large format plane arrays (FPAs). However, today's HSI systems are limited to scanning mode detection, and are usually large, complex, power hungry and slow. The ability to perform HSI in a staring mode is critical to real-time targeting mission. This brings up conflicting requirements for real-time, large area search with the ability to detect and identify difficult and hidden targets using hyperspectral information, while staying within the processing time and size available suited to small platforms. DARPA's Adaptive Focal Plane Array (AFPA) Program is to develop an electro-optical imaging sensor that benefits from both hyperspectral and FLIR, while avoiding the large mass of hyperspectral and the poor target-to-background signal differential. The AFPA designs allow a pixel-by-pixel wavelength selection in hyper-spectral imaging. With continuous spectral tuning, users may re-program the AFPA based upon the characteristics of the target and background. However, this requires the ability to decide which wavelengths to select. As researchers have not identified any unique band of interest, making the sensor spectral tunable is not sufficient. It may be that measurements should be made at collection time that enables the bands to be selected and the associated algorithms tuned correspondingly. In addition, the limited field of view (FOV) of conventional sensor design does

not satisfy the requirements of large area search. With this in mind, instead of starting with what technologies can provide, we start with a single bigger question: what do users really need? Under this new paradigm, the three major sub-questions that we ask in a "performance-driven" context are as the following.

1. **System evaluation:** Given a real-world task, how could we rapidly prototype, optimally utilize and evaluate a multimodal sensor, using a general framework and a set of modeling tools that can perform a thorough and close-loop evaluation of the sensor design?
2. **Sensor description:** Given a real-world task, what are the optimal sensing configurations, subsets of data and data representation that are most decision-relevant to provide guidelines for adaptive multimodal sensor designs?
3. **Data exploitation:** Given a real-world task, what advanced data processing and exploitation are needed to support intelligent data collection?

Base on these we propose a system approach for adaptive sensor designs that is possible to reduce development time and system cost while achieving better results through an iterative process. With this approach, it is possible to reduce development time and system cost while achieving better results through an iterative process that incorporates user requirements, data and sensor simulation, data exploitation, system evaluation and refinement.

1.2. A system approach

Conventionally, the development of a multimodal sensor system requires that many components be selected and integrated in a manner that fits a task and maximizes performance. Such system includes a variety of design tradeoffs that would be difficult and expensive to determine by building physical prototypes. It is inflexible because of the difficulty in changing early design decisions when that would imply more investigations and trade studies. Furthermore, it is difficult to include the end users in the process and to thoroughly evaluate the sensor performance.

The need of a generic system design can reduce the development time and cost by modeling the components and simulating their response using synthetically generated data. This is implemented through scene and sensor simulation tools to model and simulate the background and target phenomenology and sensor characteristics, and place them in a realistic operational geometry. The proposed framework is based on the Digital Imaging and Remote Sensing Image Generation (DIRSIG) (DIRSIG 2008) tools for characterizing targets, environments and multimodal sensors. Through a realistic scene simulation and sensor modeling process, ground truth data are available for evaluating the designed sensors and related vision algorithms. The simulation tools also allow us to more effectively refine our sensor designs. A Data Process Management Architecture (DPMA) is designed, which is a software system that provides a team

development environment and a structured operational platform for systems that require many interrelated and coordinated steps.

As a case study, we use a peripheral-fovea design as an example to show how the evaluation and refinement can be done within a system context. This design is inspired by the biological vision systems for achieving real-time imaging with a hyperspectral/range fovea and panoramic peripheral view. Issues of sensor designs, peripheral background modeling, and target signature acquisition will be addressed. This design and the related data exploitation algorithms will be simulated and evaluated in our general data simulation framework.

Under the principle of the system approach and performance-driven sensing, and in the spirit of the peripheral-fovea design, a real multimodal human signature detection sensing design is proposed and a test prototype is developed to capture visual, audio and range information at a large distance. The core functionality of the system is remote hearing using a unique optical sensor – Laser Doppler Vibrometer (LDV). Further, a video camera is used to get visual information of the target and finds the right objects for LDV to hear. In addition the camera together with the LDV measure the distance of an object/subject for the purpose of LDV focusing and object range estimation. In this example, the PTZ camera serves as the peripheral vision, and the fovea sensing have unique acoustic and depth measurements for target identification.

This report is organized as the following. Section 2 illustrates the system design framework. Section 3 shows the design of the bio-inspired adaptive multimodal sensor platform - the dual panoramic scanners with hyperspectral/range fovea (DPSHRF) for the task of tracking moving targets in real time. Section 4 describes the simulation environment for implementing our system approach, and the parameter configuration of the sensor platform. Section 5 presents the image exploitation algorithms for detecting and tracking moving targets, and the spectral classification method in recognizing moving objects. Section 6 discusses a multimodal sensing platform using real-world sensors. Conclusions and discussions will be provided in Section 7.

2. System Architecture

The Data Process Management Architecture (DPMA) is a software system that has been under development in an evolutionary manner for the last several years to support data collection, data management and analysis tasks. An early design system was called the Data Cycle System (DCS) for NASA's Stratospheric Observatory For Infra-red Astronomy (SOFIA) (Becklin et al. 1998). The DCS provides the primary interface of this observatory to the science community and supports proposing, observation planning and collection, data analysis, archiving and dissemination. The RIT Laboratory for Imaging Algorithms and Systems (LIAS) is the lead in

the DCS development under contract to the Universities Space Research Association (USRA) and works with team members from UCLA, University of Chicago, NASA ARC and NASA GSFC. A second system was developed and is in operational use to support real-time instrument control, data processing and air-to-ground communications as a part of the Wildfire Airborne Sensor Program (WASP) project. The goal of WASP is to provide a prototype system for the US Forest Service to use in wildfire management. The real-time component for WASP, called the Airborne Data Processor (ADP), was constructed using the knowledge we had gained in doing the DCS project, but it is significantly different. Its real-time processing supports geographic referencing and orthographic projection onto standard maps (*e.g.,* WGS-84), mosaic generation, and detection of events and targets of interest.

The DPMA is a design that is based on the experience with both the SOFIA DCS and WASP ADP as well as other activities related to distributed processing, archiving, computing and collaborative decision support. It provides: 1) An adaptable workflow system, capable of managing many simultaneous processing tasks on large collections of data; 2) A set of key abstractions that allow it to be agnostic regarding both data formats as well as processing tasks.

While analyzing the requirements of a system supporting the long term archival and workflow requirements of sensing and image processing systems, we identified three key abstractions. These are the core elements of the DPMA, and are the basis on which we can offer a system that is flexible in its support of algorithms, scalable in its workload, and adaptive to future growth and usage.

The first element of our architecture is the management of the data itself. Starting with an archive supporting a wide variety of data types (e.g., images, vectors, and shape files), new data types can be added to the system by writing additional front ends to this archive as needed. Data stored in this archive can be available for future processing or exchange with other image processing professionals, and mined in the future to generate indexes as new features become relevant. Most importantly, data in this archive can be grouped together into collections to be processed by various agents and operators; any instance of some particular data may be named in multiple collections, supporting multiple and simultaneous assignments of work in the DPMA. Finally, new data and new versions of existing data are accepted by the archive, but the data it replaces is not lost; in this way, we support historical accuracy and analysis, as well as quality control and evaluation of competing algorithms over time.

The second element of our architecture is the specification of processing agents. Each step in an image processing chain, be it an implementation of an automated algorithm, an interactive tool driven by an imaging expert, or even a simple quality assessment by reviewers, is embodied as a processing agent. We provide a support layer for these agents that provide a mechanism for delivering the materials in a work assignment from the archive to the agent, as well as a similar

mechanism for storing all new data products yielded by an agent back in the archive, making them available for another agent. This support of processing agents allows the DPMA to evolve image processing and analysis tasks from those that may be performed by a single operator at an image processing workstation to clusters or specialized hardware implementing developed and groomed algorithms in an automated and unattended fashion.

The third element of our architecture is the definition of the workflows themselves. A workflow will be a directed graph whose nodes are the aforementioned processing agents. In the archive, a work assignment is associated with a workflow, moving through the nodes of its graph to reflect its current processing state. Because an agent implementing a processing step can just as easily be a quality assessment, or ”gate”, as it can be an image processing or decision algorithm, it is easy to instrument workflows with progress reviews, data evaluations, and so on. Similarly, just by manipulating the state of a work assignment (that is, by moving an assignment to a different node in a workflow graph), it is trivial to repeat previous steps in the workflow with different algorithm parameters or operator instructions until a reviewer is satisfied that the assignment can proceed to the next step in the workflow.

These three resources: bundles of data as a work assignment, intelligent distributed agents, and processing workflows, are orthogonal to each other. They all reference each other, but they are defined independently and separately. Each contains entities that name or reflect entities of the other two elements. Taken separately, each part of the system can be grown independently over time with improvements to existing entities or entirely new entities; this growth does not affect entities elsewhere in the system, and dramatically reduces the typical overall risk of system upgrades. Taken together, we have a system that can be easily adapted to new types of data (archive), new processing steps (agents), or new approaches to solving a problem (workflows).

To further demonstrate our system architecture for managing data process we will first describe a complex sensor design that effectively uses a small hyperspectral fovea to gather only important data information over a large area.

3. A Bio-Inspired Sensor Design

To break the dilemma between FOV and spatial/spectral resolution for applications such as wide-area surveillance, we investigate a bio-inspired data collection strategy, which can achieve real-time imaging with a hyperspectral/range fovea and panoramic peripheral view. This is an extension of the functions of human eyes that have high-resolution color vision in the fovea and black-white, low-resolution target detection in the wide field-of-view peripheral vision. The extension and other aspects of our system are also inspired by other biological sensing systems (Land & Nilsson, 2004). For instance, certain marine crustaceans (e.g., shrimp) use hyperspectral

vision in a specialized way. In our system, the hyperspectral vision is only for the foveated component. As another example, each of the two eyes of a chameleon searches 360-degree FOV independently. This inspires us to design two separate panoramic peripheral vision components. Some species (such as bats and dolphins) have excellent range sensing capabilities. We add range sensing in our simulated fovea component, and later on we will also extend to foveated system with a laser Doppler sensor to measure acoustic signals at a large distance (Section 6).

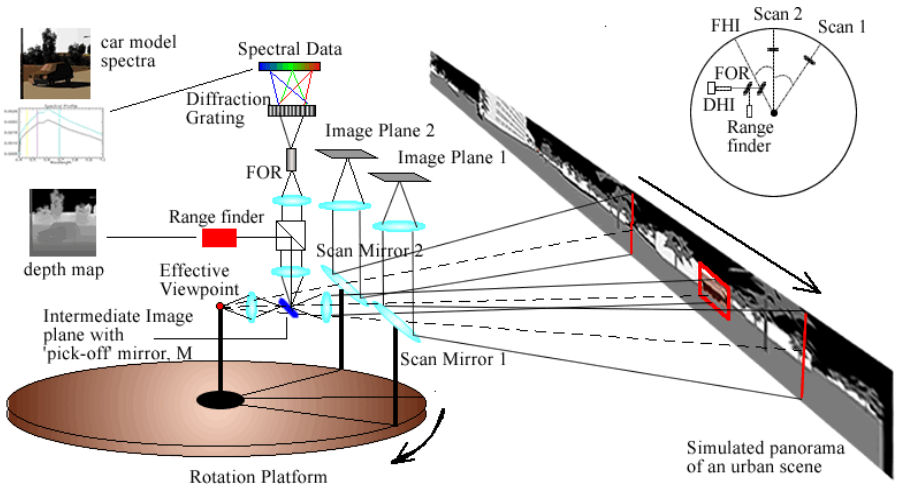


Figure 1. The design concept of the DPSHRF. The dash lines indicate the single viewpoint of both the foveal hyperspectral imager and the two line scanners.

The data volumes in consideration have two spatial dimensions (X and Y), a spectral dimension (S, from a few to several hundred), and a time dimension (T). This four dimensional (4D) image in X-Y-S-T may be augmented by a 2D range image (in the XY space). Ideally, a sensor should have 360-degree full spherical coverage, with high spatial and temporal resolution, and at each pixel have full range of spectral and range information. However, this type of sensor is difficult to implement because of the enormous amount of data that must be captured and transmitted, most of which will eventually be discarded. Therefore, particularly for real-time applications, every collection must face fundamental trade-offs such as spatial resolution vs. spectral resolution, collection rate vs. SNR, field-of-view vs. coverage, to name a few examples.

Understanding the trade-offs and using algorithms that can be adapted to changing requirements can improve performance by enabling the collection to be done with maximum effectiveness for the current task. In our design, the fovea is enhanced by HSI and range information, and the peripheral vision is extended to panoramic FOV and has adaptive spectral response rather than just black-white.

Our proposed sensor platform, the dual-panoramic scanners with a hyperspectral/range fovea (DPSHRF) (Fig. 1), consists of a dual-panoramic (omnidirectional) peripheral vision and a narrow FOV hyperspectral fovea with a range finder. This intelligent sensor works as the follows: In the first step, two panchromatic images with 360-degree FOV are generated by rotating two line scanners around a common rotating axis, pointing apart to two slightly different directions. The angle difference between the two scanners can be adjusted for detecting and tracking moving targets with different velocities and distances. An initial angle is used at the beginning. Then the detecting results from the two scans can determine what the new angle difference should be - either decreased if a target is moving too fast, or increased if the target is moving too slow. There are two advantages of using line scanners that will be further amplified. First, a line scanner can have a full 360-degree horizontal FOV. Second, resulted images are inherently registered.

Moving targets can then be easily and quickly determined by the differences of the two panoramic images generated from two scanners. The next position and the time of a moving target can be estimated from the difference of two regions of interest (ROIs) that include the target. In real-time processing, the comparison is started whenever the second scan reaches the position of the first scan, therefore, only a small portion of panoramic images is used before full-view panoramas are generated. The detail of the target detection processing algorithm will be discussed in Section 5.

Then, we can turn the hyperspectral/range fovea with a specific focal length calculated based on the size of the object, and to the predicted region that includes the moving target. Thus, hyperspectral/range data is recorded more efficiently for only the ROIs that include possible moving targets. The two line scanners and the hyperspectral/range imager are aligned so that they all share a single effective viewpoint. The spectral data can be efficiently recorded with a foveal hyperspectral imager (FHI) (Fletcher-Holmes and Harvey, 2005) which maps a 2D spatial image into a spatial 1D image. This is implemented by using a micro mirror as a fovea that intercepts the light onto a beam splitter for generating co-registered range-hyperspectral images using a ranger finder and the FHI. The FHI consists of a fiber optical reformatter (FOR) (Fibreoptic) forms a 1D array onto a dispersive hyperspectral imager (DHI) (Headwall) which produces a 2D hyperspectral data array with one dimension as spatial and the other as spectral. The spatial resolution of the FOR is determined by the diameters of optical fibers which are controlled during the optical design process. The blurring effect from cross-coupling of optical fibers is not significant magnitude as shown in (Harvey and Fletcher-Holmes 2002). Finally, a co-registered spatial-spectral/range image is produced by combining with the panchromatic images which are generated by the dual-panoramic scanners.

In summary, this sensor platform improves or differs from previous designs (Goldberg et al. 2003, Fletcher-Holmes and Harvey 2005, Harvey and Fletcher-Holmes 2002) in literature in four aspects:

1. A dual scanning system is designed to obtain moving targets in a very effective and efficient manner. A panoramic view is provided instead of a normal wide-angle view.
2. An integration of range and hyperspectral fovea component is used for target identification.
3. The dual-panoramic scanners and the hyperspectral/range fovea are co-registered.
4. Active control of the hyperspectral sensor is added to facilitate signature acquisition of targets of various locations that can only be determined in real-time.

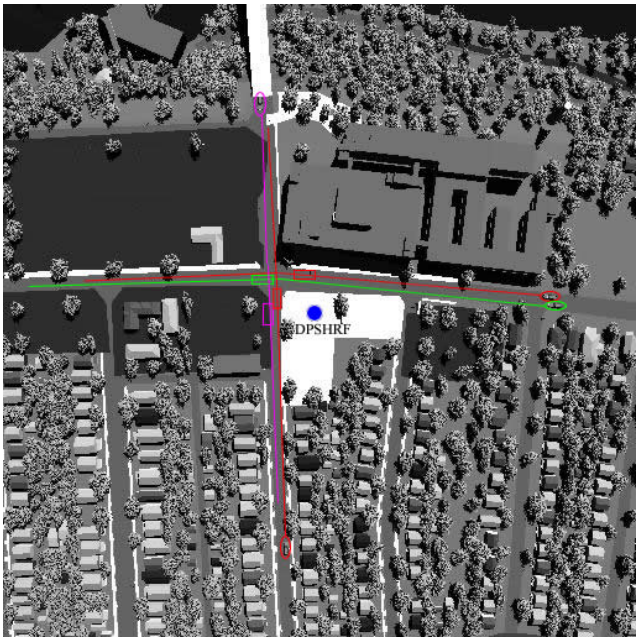


Figure 2. A simulated urban scene image captured at latitude=43.0° and longitude=77.0°, 1000 meters above. The ellipses show the initial state of the four cars. The rectangles show the state where those cars move after 12s. The "DPSHRF" sensor (in blue dot) is placed in front of the main large building. The simulated scene is captured at 8am in a typical summer day.

4. Scene Simulation and Sensor Modeling

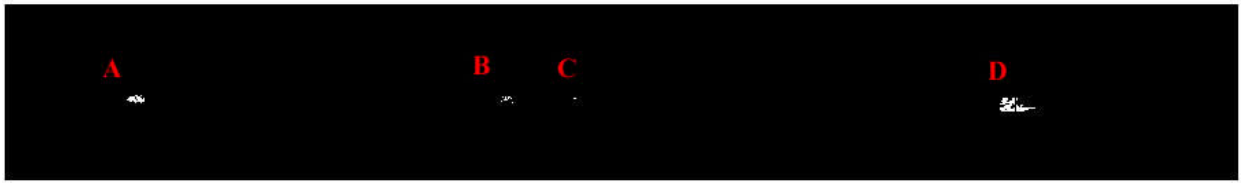
The sensor design concept is tested though the simulation tool DIRSIG. Various broad-band, multi-spectral and hyperspectral imagery are generated through the integration of a suite of first principles based radiation propagation sub-models (Schott et al. 1999). Before performing scene simulation and sensor modeling, we need to set up different scenarios and configure the sensor parameters. One of the complex scenarios we constructed including four cars having exactly the same shape and three different paints moving to different directions with various speeds (Fig. 2). All four cars will pass through the cross section at the bottom corner of the main building in the

scene at a certain time. Various behaviors of the moving vehicles such as simple moving, overtaking, passing through, and etc., are monitored by our sensor platform which is placed in front of the main building. The scan speed of each line scanner can be set from 60 Hz to 100 Hz selectable, thus one entire 360° scan take from 6.0 seconds down to 3.6 seconds. This time constraint is not a problem for real-time target detection since detection and scanning are continuous and simultaneous. The number of pixels per line in the vertical direction is set to 512 to match the horizontal scanning resolution. Few selected spectral bands are captured by dual line scanning. The focal length is fixed at 35mm for both line scanners, and the angle between the pointing directions of the two scanners is 10° so that the time the second scan reaches the position of the first scan is only about 0.1s. In theory, the time difference between two scans should be much less than one second to avoid a lot of uncertainty of action changes in moving vehicles. Two scanners are used so that (1) the more accurate direction and the focal length of the hyperspectral fovea can be estimated; and (2) moving target detection can still be performed when background subtraction using a single scanner fails due to cluttered background, multiple moving targets, and the ego-motion of the sensor platform. The focal length of the hyperspectral imager is automatically adjusted according to the target detection results generated from the two line scanners. To simulate the hyperspectral imager, we use a frame array sensor with small spatial resolution at 70 x 70 for the hyperspectral data, and the ground truth range data provided by DIRSIG are transformed into range images. The spectral resolution is 0.01 μm ranged from 0.4 μm to 1.0 μm. Different portion of bandwidth can be selected and determined by analyzing the model spectral profile.

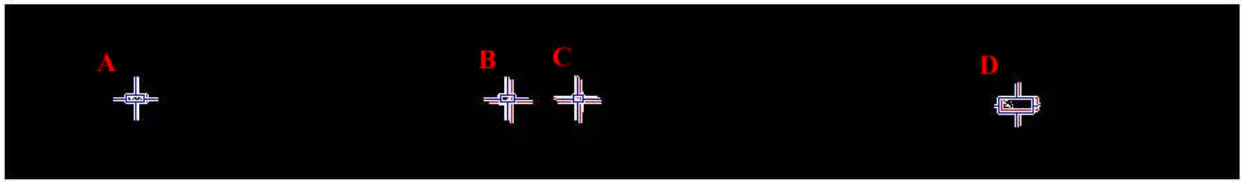
The simulation will enable a close investigation of intelligent sensor designs and hyperspectral data selection and exploitation for user designated targets. The DIRSIG simulation environment allows us to use an iterative approach to multimodal sensor designs. Starting with user and application requirements, various targets of interest in different, cluttered background can be simulated using the scene-target simulation tools in the DIRSIG. Then the adaptive multimodal sensor that has been designed can be modeled using the sensor modeling tools within the DIRSIG, and multimodal sensing data (images) can be generated. Target detection/identification, background modeling and multimodal fusion algorithms will be run on these simulated images to evaluate the overall performance of the automated target recognition, and to investigate the effectiveness of the initial multimodal sensor design. The evaluations of the recognition results against the given “ground-truth” data (by simulation) can provide further indicators for improving the initial sensor design, for example, spatial resolution, temporal sampling rates, spectral band selection, the role of range information and polarization, etc.. Finally, a refined sensor design can again be modeled within the DIRSIG to start another iteration of sensor and system evaluation.

a). Panoramic image from the first scanner, with the moving targets indicated inside red circles.

b). Panoramic image from the second scanner, again the same moving targets indicated inside green circles.



c). Frame difference between b and c, group of ROIs are labeled.



d). Background subtraction from two scans inside boundaries defined by c. Red rectangles showed ROIs from first scan, blue rectangles showed ROIs from second scan. (Close-up view of each labeled region can be seen clearly in Table 1).

Figure 3. All 360° panoramic images (512 x 3600) shown here are integrated from vertical scan lines captured by the dual-panoramic scanners.

5. Data Exploitation and Adaptive Sensing

The basic procedure for active target detection and tracking is as the follows. A few selected spectral bands are used to initialize the detection of targets either based on motion detection or scene/target properties in prior scenarios. Then, for the potential interesting targets, the fovea turns to each of them to get a high-resolution, hyperspectral image with range information. This can be done in real-time so that tracking of one target and switching between multiple candidates is made possible. Finally, the signatures of the targets can be obtained by automatically

analyzing the hyperspectral data in the fovea and by selecting the most relevant bands for such targets. This kind of function needs the active control of the sensor to fuse the peripheral and fovea vision in an efficient manner. In the following, we elaborate the principle by using some commonly used algorithms in target detection, tracking and identification, using our bio-inspired multimodal sensor.

5.1. Detection and tracking in peripheral views

The first step is to find ROIs that possibly contain moving targets (Fig. 3). Simple background subtraction between a scanned image and a background image is not sufficient because the panoramic background (with trees, building, etc.) may change due to illumination changes over a large span of time. The advantage of using the two consecutive scanners is the ability to quickly detect a moving target in real time using “frame difference” without producing too much noise from the background. Further, a morphological noise removal technique (Soille 1999) is applied to remove small sparse noises with the opening operation and fill small holes with the closing operation. However, the results from “frame difference” cannot provide accurate location and size information of the moving targets. Therefore, bounding boxes are defined from the “frame difference” results to mask off those background regions for background subtraction, which can provide more accurate location and size information of the moving targets. Fig. 3c shows some bounding boxes that can be used as masks for performing the background subtraction of each individual panoramic scan. The threshold is set very low since we are interested in any changes in motion comparing to the relative static background. Of course, false alarms can also be generated by events such as the change of a large shadow, but this can be verified once we captured the hyperspectral image. The background image is updated for only those pixels belonging to the background after each 360-degree rotation, thus moving object extraction is maintained over time. At every rotation, each of the two line scanners will generate a sequence of 1D image lines that are combined to generate the panorama. Thus, registration problems can be avoided with the stabilized line scanners. Real-time target detection can be achieved since the scanning and detection are performed simultaneously and continuously.

The next step is to estimate the region of the next position that may contain a target once the two ROIs of the same target are found at two different times resulting from two different scans (Fig. 3d). The location and size differences of the two regions can determine the relative bearing angle of the hyperspectral/range fovea imager to zoom on the moving target. The position of extracted region from the dual-scans indicates which direction the target is moving to. Also, the size of two regions can indicate whether the target is moving closer to the sensor or farther. Therefore, we

can calculate the next position where the target will be. Then, the ratio of the previous two regions can be used to estimate the new focal length of the hyperspectral imager.

The angle difference of two scans for two ROIs at different times t_i and t_{i+1}, can be used to predict the position of the next ROI having the moving target at the time, t_{i+2}, when the hyperspectral/range imager can be in place. Therefore, given the time, we can estimate the panning and tile angles of the hyperspectral/range imager. Note that only the angles relative to the center of a region are needed. The turning angles (i.e., panning and tilting) of the hyperspectral/range imager should be:

$$\theta_{t_{i+2}}^{(x,y)} = \theta_{t_{i+1}}^{(x,y)} + \frac{t_{i+2}^{(x,y)} - t_{i+1}^{(x,y)}}{t_{i+1}^{(x,y)} - t_i^{(x,y)}} (\theta_{t_{i+1}}^{(x,y)} - \theta_{t_i}^{(x,y)}) \qquad (1)$$

where the superscript x and y correspond to the panning angle (in the x-direction) and the tilting angle (in the y-direction), respectively. The angle θ_{t_i} corresponds to the angle position of a ROI at a time t_i as shown in Fig. 4. The focal length of the hyperspectral/range fovea is inversely proportional to the desired FOV of the hyperspectral/range imager, α, in order to have the target in the full view of the FOV. The FOV angle can be estimated as

$$\alpha = \frac{R_{t_{i+2}}}{P^l} \qquad (2)$$

where $R_{t_{i+2}}$ is the predicted size of the target region at t_{i+2}, and P^l is the number of scanning lines per radius. The relationship between $R_{t_{i+2}}$ and the previous two regions of the same target at different times can be expressed as

$$\frac{R_{t_{i+1}}}{R_{t_i}}(t_{i+1} - t_i) = \frac{R_{t_{i+2}}}{R_{t_{i+1}}}(t_{i+2} - t_{i+1}) \qquad (3)$$

Then a hyperspectral foveal shot of a ROI from the calculation can be taken. Thus, hyperspectral/range data is recorded in a more efficient way, only for ROIs. It is possible for some regions to be identified that do not have true moving targets inside. Then the hyperspectral classification in next step can verify this situation.

5.2. Target classification using 3D and HSI fovea

Targets can be classified based on hyperspectral measurements, shape information, and the integration of both. There has been a lot of work in recognizing objects using 3D shape information (e.g., Diplaros et al. 2006; Start and Fischler 1991). Here we will only describe how to use a target's depth information and the information of its background to perform better hyperspectral classification.

Recognizing a target needs to compare the target's spectrum associated with each pixel to its training spectrum. In our experiments, a spectral library was pre-built with some existing models. Various vehicles with different colors and shapes can be imported and tested in the simulation scene. In the particular scenario in Fig. 2, four cars having the same shape but different paints are modeled. Two are red, one is brown and one is black. Initial spectral signatures of the four cars were captured from different angles in the same background. The capturing angles and surroundings are important and need to be considered carefully because those factors can significantly affect the effective radiance reaching the sensor, $L(l,\theta,\varphi,\lambda)$, where l is the slant range from sensor to target, θ, φ and λ are the zenith angle, the azimuth angle and the wavelength, respectively. The general expression for L is more complex and fully described in (Schott, 2007). However, we can simplify L if we are only interested in the reflective (visible) bands, the general equation can be further expressed as:

$$L(l,\theta,\phi,\lambda) = f\left(L_s(l,\sigma,\lambda),\ L_{ds}(\theta,\phi),\ L_{bs}(\theta,\phi,\lambda), L_{us}(l,\theta,\lambda)\right) \qquad (4)$$

where σ is the angle from the normal to the target to the sun, L_s is the solar radiance, L_{ds} is the downwelled radiance from the sky due to the atmospheric scattering, L_{bs} is the spectral radiance due to the reflection from background objects, and L_{us} is the scattered atmospheric path radiance along the target-sensor line of site.

In the training stage, the background is known and fixed, thus L_{bs} can be cancelled out. The angles of the sun to the target and the of target to the sensor are known, thus we can keep this information and estimate a new spectral profile of the model target once we need to monitor a new target at a different time. L_{ds} and L_{us} can also affect the initial spectral profile if the weather condition changes significantly. In the current experiments, we only use one atmospheric dataset which can also be replaced and changed in the simulation in the future. After handling all reflective variants, various endmembers that represent the spectral extremes that best characterize a material type of a target were selected, and their spectral curves were stored in the spectral library database. We used the sequential maximum angle convex cone (SMACC) (Gruninger et al. 2004) to extract spectral endmembers and their abundance for every model target. In comparison to the conventional pixel purity index (PPI) (Boardman et al. 1995) and N-

FINDER (Winter 1999), SMACC is a much faster and more automated method for finding spectral endmembers. Simply speaking, SMACC first finds extreme points or vectors that cannot be represented by a positive linear combination of other vectors in the data as a convex cone, and then a constrained oblique projection is applied to the existing cone to derive the next endmembers. The process is repeated until a tolerance value is reached, for example, max number of endmembers. Each endmember spectrum, defined as *H*, can be presented mathematically as a combination of the product of a convex 2D matrix contains endmember spectra as columns and a positive coefficient matrix:

$$H(c,i) = \sum_{k}^{N} R(c,k)A(k,j) \qquad (5)$$

where *i* is the pixel index, *j* and *k* are the endmember indices, and *c* is the spectra channel index. Some endmembers might have less spectra differences in term of redundancy. Those can be coalesced based on a threshold so that the most extreme spectra are identified and used to represent the entire coalesced group of endmembers.

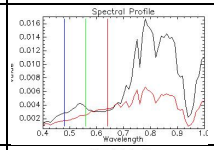
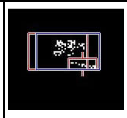
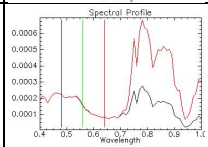
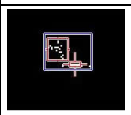
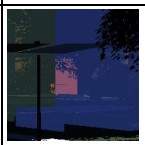
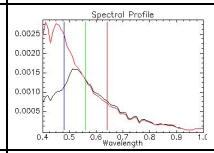
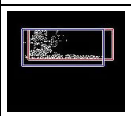
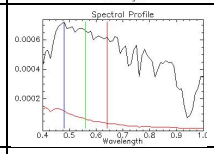
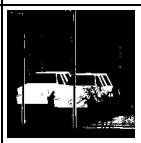
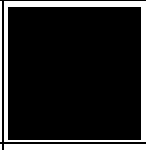
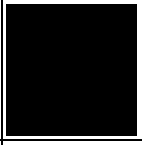
In the testing stage, the same target spectra may be varied in different conditions such as various surface orientations and surroundings. However, the significant spectral signature of a target can be estimated and maybe further corrected with the help of range information produced from a ranger finder. Knowing the angles of the sun and the sensor, the depth map (i.e. range data) can indicate whether the information of a background object close to the target should be counted when processing the target spectra. The result spectra will have similar shape but the magnitudes will be still different due to the variations of illumination intensities and directions. A spectral angle mapper (SAM) (Kruse et al. 1993) algorithm is used to match the target spectra to reference spectra. The SAM is insensitive to illumination and albedo effects. The algorithm determines the spectral similarity between two spectra by calculating the angle between the spectra and treating them as vectors in a space with dimensionality equal to the number of bands (Kruse et al. 1993). Smaller angles represent closer match. The depth information and the relative location of the sun and the sensor can determine whether a target spectra should be adjusted by the surrounding spectra when performing classification. As a result, each pixel is classified either to a known object if the target spectrum is matched with the library spectrum of that object, or to an unknown object, for instance, the background. To distinct multiple objects from database, the results from different group of endmembers of different targets are compared.

5.3. Experimental results

Table 1 shows the processed results for the following four cases: A) Multiple targets with different spectral signature. B) A target is under a shadow cast by trees. C) There is no moving target (thus a false alarm). D) Only one side of the target spectral signature can be acquired and the other side cannot be determined due to the insufficient reflectance of the sun light and the surroundings. At this stage, we only recognize if the detected target is the car or not the car. The target region may not fully match to the right shape of the car model. Only the sub-region with sample pixels spectra are selected for the matching. From scenarios A and B, both frontal and side shape of the car can be recognized. However, in scenario D, the side of the car cannot be detected due the shadow from the nearby building. We also captured multiple shots when those cars moved to various locations following the trajectories indicated in Figure 2. We can recognize those cars with different colors, but most false targets are resulted from the large shadow. Various solutions can be possible, for example: 1) to place the sensor platform at another position; 2) to reconfigure sensor parameters such as adjust the height and the pointing direction; and 3) to implement a better classification algorithm. Therefore the experimental results can quickly drive feedback to adjust and improve the sensor design and the algorithm implementations. Various scenarios and cases can be constructed and tested in the simulation framework before a real sensor is even made.

One of the useful advantages of the co-registered hyperspectral and range imaging is to using the range information to improve the effectiveness of the hyperspectral measurements. For example, in Table 1B, the shadowing of the vehicle (the red car) under the trees can be analyzed by the relation among the location of the sun, the locations of the trees from the panoramic background, and the surface orientations of the vehicle. Considering the depth information, the SAM can be obtained for surfaces of the vehicle under the influence of the tree shadows (therefore looks greenish). In Table 1D, the color only information is not sufficient to recognize the right target at where the background is also selected as the same one. With the depth information, the relations between the surfaces orientations of the vehicle (the black car) and the location of the sun can also tell which surfaces are illuminated. Therefore the well-illuminated surfaces (i.e. the top of the car body) can be selected based on the structural information obtained from the range data. The analysis so far is very preliminary but is very promising for future research.

Table 1. Processing Results of the simulated urban scene.

Index	ROIs	Fovea Parameters	Fovea Shot	Sample Spectral Profile	Spectral Curves Annotations
A		Zenith: 89.0 Azimuth: 80.0 Focal Length: 245mm			<u>Top Curve:</u> Car on Right <u>Bottom Curve:</u> Car on Left
B		Zenith: 88.5 Azimuth 191.0 Focal Length: 205mm			<u>Top Curve:</u> Car not in Shadow <u>Bottom Curve:</u> Car in Shadow
C		Zenith: 88.5 Azimuth 220.0 Focal Length: 225mm			<u>Top Curve:</u> Material 1 <u>Bottom Curve:</u> Material 2
D		Zenith: 88.0 Azimuth 330.0 Focal Length: 125mm			<u>Top Curve:</u> Front body <u>Bottom Curve:</u> Side body
Index		Depth Map	SAM no depth	SAM with depth	Results
A					Red Car Brown Car
B					Red Car
C					False Target
D					Black Car

Each index corresponds to each labeled region in Fig 2d. The column ROIs shows close-up view of result indicated in Fig 2d. Hyperspectral fovea shots demonstrated here with only 3 RGB bands (which are also marked as vertical lines in the sample spectral profile column, in blue, green and red, respectively. Only the significant spectral signatures of targets are shown here. Final mapping results are shown in binary only to indicate the targets and the background. The classification is based on the match result with each model target spectral profile in database.

6. A Multimodal Sensing Platform with Real Sensors

Under the principle of the system approach and performance-driven sensing, a real multimodal human signature detection sensing design is proposed and a test prototype is developed to capture visual, audio and range information at a large distance (Qu, et al, 2009; Qu, et al, 2010, Wang, et al, 2010). The core functionality of the system is remote hearing using a unique optical sensor – Laser Doppler Vibrometer (LDV). Further, a video camera is used to get visual information of the target and finds the right objects for LDV to hear. In addition the camera together with the LDV measure the distance of an object/subject for the purpose of LDV focusing and object range estimation. This is a real-sensor example of our generalized multimodal peripheral-fovea sensing designs. In this case, the PTZ camera serves as an active peripheral vision component for target detection, whereas the fovea component is further extended to have both range and acoustic measurement capabilities. Since it is an on-going work starting with the AFOSR grant, we will only give a brief overview here; we would like to continue to study more fundamental issues in adaptive multimodal acquisition and understanding for long-range surveillance in our future research.

6.1. Problems

Laser vibrometry has found a lot of important applications in medical, industrial, surveillance and inspection fields. However, most of the systems are manually operated. In close-range and lab environments, this is not a serious problem. But in field applications, such as bridge/building inspection, area protection or battle field applications, the manual process takes very long time to find an appropriate reflective surface, focus the laser beam and get vibration signal. For example, it is very hard to aim the laser beam to a reflective surface if it is 100 meters away. Even if the laser beam is pointed to the surface, it takes quite some time to focus the laser beam; but there is no guarantee that the signal return will include the vibration signals needed. As an example, using the Polytec 505 LDV, the built-in automatic focusing takes about 15 seconds to focus the laser beam on the surface of a target. Finally, in some monitoring application, it is desired to have the LDV system being adaptive to the changing environments, such as acquiring acoustic (voice) signals for subjects (humans or vehicles) entering a protected perimeter. Therefore there are great unmet needs in facilitate the process of surface detection, laser aiming, laser focusing, and signal acquisition of the emerging LDV sensor, preferable through system automation. Here we apply our adaptive multimodal sensor design principles to this unique sensor and develop it into an exploitation-driven multimodal sensor system, the vision-aided automated vibrometry (VaaV).

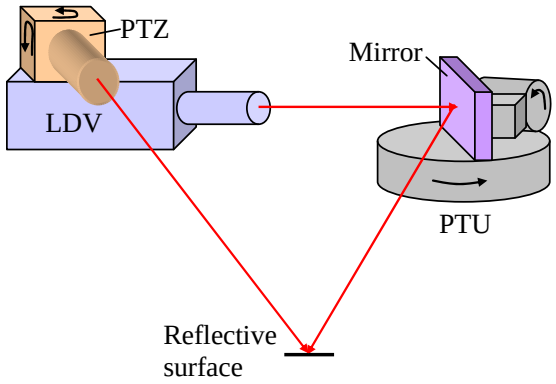


Figure 4. A multimodal sensing platform for long-range surveillance

The vision-aided automated vibrometry (VaaV) includes an automated system that is added upon an existing device: laser Doppler vibrometer. The automated system includes the following hardware components: a pan-tilt-zoom (PTZ) camera, and planar mirror M on a pan-tilt unit (PTU), and a personal computer with the appropriate interfaces to the devices (USB, RS232, Firewire, etc). The system also includes a software component with algorithms to analyze the signals and control the devices. Figure 1 show the system diagram.

6.2. Vision-aided surface detection, orientation estimation and laser aiming and focusing

The main function of the vision-aided vibrometry is an automatic laser focusing function. It uses the information about the range of the reflective surface, and the signal levels (strengths) of the LDV signal returns to automatically focus the laser beam to the target's surface. As we have noted, some of the LDV systems have automatic focusing functions, but usually they take quite long time, e.g. 15 seconds for Polytec OFV-505. With our automatic focusing method, our current implementation reduced the time to 1.5 second, a ten-fold improvement (Qu, et al, 2010).

In un-controlled environment, another major issue is to find an optimal surface that the LDV can best detect the desired vibration signals. This requires the surface vibrate well with the vibration source and well reflect the LDV laser beam. For example, for bridge and building inspection, we would like to find an appropriate surface (e.g. façade of the building) that has both reflection and vibration. For acquiring human voice signal from a large distance, we have found it is very difficult to achieve this by pointing the laser to the person. We will have to find a reflective surface that well vibrates with the human voice. Furthermore, for a large distance, it is very hard to see the laser spot by bare eyes.

Our vision-aided vibrometry approach eases this problem greatly. Since the PTZ camera and the LDV system (with the reflecting mirror on a PTU) are calibrated, then we use the PTZ camera to both find an appropriate reflecting surface and to control the PTU to aim the laser on the surface. This active sensing approach enables both interactive surface selection and laser pointing, and

full automation of these two processes (Qu, et al, 2010). Therefore it is useful for both applications and research. As soon as the laser points to the selected surface, the automatic distance measurement and laser focusing will be performed in less than two seconds.

The vision-aided automated vibrometry can also estimate the surface orientation. After the distance estimation of more than three points, a planar or high order surface model can be fitted to the data. The surface orientation information is useful not only for assisting the evaluation of the reflecting signals, but also for understanding the geometry of the surface.

The VaaV system can be further used as a two-dimensional (2D) LDV scanning system. By pan and tilt the mirror on the PTU, vibration signals can be obtained on a grid of points on the selected surface. The VaaV system will generate multimodal data for the surface: a 2D range map encoded the distances of the selected points on the surface, the 2D vibration map, and a color image of the same surface, all perfectly aligned. The multimodal information will be very useful for analyzing the material, geometric and dynamic properties of the targets, such as in building/bridge/vehicle inspection.

6.3. Automatic and adaptive audio-visual signal acquisition and processing

This work also has a unique function of automatic voice signal acquisition and processing. After the laser beam is focused on the target, the signal is automatically collected, analyzed, played and visualized. This will help the user or the application system to determine if the signals are the desire vibration signals. This novel feature of our work is particularly useful for applications of human signature detection in remote surveillance, area protection, battle field living assessment, and intelligence. The voice signal processing module includes both band-pass filtering and Wiener filtering. The desired signals, as voice streams, can also be played as audio clips and be further used for human identification, speech recognition and language identification. The voice signals are further analyzed to detect acoustic events such as human speeches, vehicle engine sounds, etc. (Tao, et al, 2010).

The VaaV system also features an automatic laser pointing adaptation. While the signal is in capture, the signal level and the signal waves are being analyzed in order to check it the return signals are still valid. For example, in a multimodal human signature detection application, the VaaV module can be integrated with a human detection module so that new reflective surface can be selected when the human subject walks away from the LDV detection range. The VaaV system can be used for this purpose by using both video analysis of the PTZ camera, and the vibration signals.

The LDV is capable of detecting the acoustic signals from various vibration surfaces, including window frames, concrete wall, traffic signs, etc.. However, a reliable acoustic background modeling technique should be constructed in order to separate the outliers from the background

"sound" that includes both the real background sound and the signals created by the electronic-optical noises of the LDV. A Gaussian Mixture Model (GMM) is used to model the feature distributions of signals (Tao, et al, 2010). Then each mixture component of a surface acoustic model is represented by a unique Gaussian mean vector and a covariance matrix. However, the GMM does not build relations among different mixture components in a surface model and those components in the other surface model may be very similar to them. In order to present the temporal dependencies of components in a surface model, we use a window-based aggregation technique for the GMM with more than one component. The basic idea is to select a sequence of overlapped windows each contains consecutive features in time series and construct a normalized histogram based to the decision of those features. In general, the decision of a feature is either one of background components or the foreground. Then the average of all constructed window-based histograms for the right background model creates a temporal pattern that can be used to compare any input signals.

As an example, we constructed audio background models from various surfaces including metal box, painted metal door, chalkboard, whiteboard, and wall. We tested the models in an indoor corridor of about 420 feet long for the long range audio-visual event detection. Foreground audio events are extracted using the corresponding background surface model. The target person in the video is detected using a standard background subtraction technique in computer vision. Results from the audio and video are combined to demonstrate the final event decision. In Figure 5, the red box shows a person behind the wall that cannot be observed in the image but there is speaking detected from the audio stream (in shaded region). The blue box shows the people detection in the camera view.

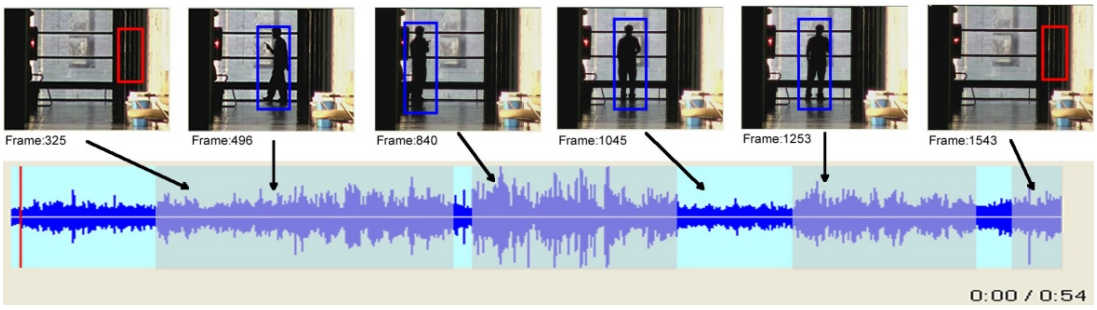


Figure 5. visual-audio integration for human detection

7. Conclusion and Discussions

We first briefly described our system architecture and its characteristics accommodating with sensors design and algorithm. Then, we mainly described our bio-inspired multimodal sensor

design that enables efficient hyperspectral data collection for tracking moving targets in real-time. This design and the related processing steps are tested through a system approach with sensor modeling, realistic scene simulation, and data exploitation. By simulation, various components can be reconfigured or replaced for specific situations or tasks. The image processing algorithms are designed only to demonstrate the basic idea of effectively capturing hyperspectral data in ROIs based on data exploitation. Needless to say, more sophisticated algorithms need to be developed for more challenging tasks. We only described one spectral classification method for recognizing the object. More precise and efficient hyperspectral classification routines may be applied. In addition, error characterizations of the hyperspectral sensing and range sensing have not been discussed. These are the standard procedures in image analysis and computer vision; our simulation approach will facilitate the simulation and evaluation of the system performance under various signal-to-noise ratios (SNRs). This remains our future work.

The real-time hyperspectral/range fovea imaging further extends the capability of human fovea vision, and unique capacities of other biological sensing systems. In the future, we will study two aspects of data processing: range-spectral integration and intelligent spectral band selection. Both issues will be greatly facilitated by our system approach and advanced scene and sensor simulation.

Range-spectral integration. There are many factors that need to be considered in correcting the acquired hyperspectral data to reveal the true material reflectance, including source illumination, scene geometry, atmospheric and sensor effects, spectral and space resolution, and etc. In the low-altitude airborne or ground imaging cases, the scene geometry is probably the most important factor. Therefore, the design of co-registered hyperspectral and range fovea will provide both spectral and geometry measurements of the 3D scene in a high resolution, so that a range-aided spectral correction can be performed. Using the DIRSIG tools, we have simulated both hyperspectral images and ranges images for several selected targets with known 3D models and spectral properties, and the next step to derive algorithms to perform spectral correction by the more effective 3D structure information of the targets given by the range images and the background information given by the panoramic scanners.

Optimal band selection. After the analysis of the hyperspectral data, the most useful wavelengths that can capture the target's signatures can be selected via tunable filtering; and the task of tracking and target recognition will only need to use the few selected bands or a few key features rather than all of the bands. This study will be carried out in several scenarios involving different targets in a challenging background or different backgrounds. We will compare the hyperspectral profiles (i.e. 3D images with two spatial dimensions and a spectral dimension) of various targets against different background materials, and then derive the optimal spectral

signatures to distinguish a target from its background. We will also investigate how the range information can be used in improving the effectiveness of signature extraction and target recognition. The DIRSIG target and scene simulation tools could provide sufficient samples as training examples for us to optimal hyperspectral band selection.

On one hand, the design of a comprehensive system architecture such as DMPA stems from the requirements of managing a large-scale, multiple tasks, and distributed sensor systems. We made the first attempt to apply a system approach with a system architecture for multimodal sensor designs. However, the results are very preliminary and it is still a challenging issue to evaluate the usefulness of such an architecture for improving sensor designs. We hope our proposed idea will stir more research interests in looking into this problem. From our very early study, any multimodal sensor design that is implemented within the DPMA framework will have a number of characteristics. First, the sensor system model will have one or more sensing devices that receive stimulation or real data from the environment. It is likely that these sensing devices operate independently and can have parameter settings controlled from system control programs. The data produced by various components should be reusable and be preserved to be compared with other results. Second, the system model will begin with a few components that monitor the behaviors of sensing and data processing, and then will grow over time as components are added and the structure is refined. All components models should be reusable. Often real sensor systems, even with multimodal functions, still have limitations. Such a system approach with both real data to be captured by real sensors and expected data to be simulated will have a benefit to study the performance of the integrated system before the new functions can be implemented. Third, simulation model can be used to describe observations which contain the state of components, the inputs and the response and actions at various times. At last, human interaction should be provided for designers not only to understand the interactions and evaluate performance of the system but also to instantiate different components, set their parameters, and, in general, prescribe all aspects of simulation.

References

[1]. Becklin, J. A. D. E. E., Casey, S. C. and Savage, M. L., “Stratospheric observatory for infrared astronomy (sofia),” in *Proceedings of SPIE* (A. Fowler, ed.), vol. 492 of *3356*, The International Society for Optical Engineering, 1998.

[2]. Boardman, J. W., Druse, F. A., and Green, R. O., “Mapping target signatures via partial unmixing of AVIRIS data.” *Fifth Annual JPL Airborne Earth Science Workshop*, Vol. 1, AVIRIS Workshop, pp. 23-26, 1995

[3]. Diplaros, A., Geves,T., and Patras, I., “Combining color and shape information for illumination-viewpoint invariant object recognition”. *Image Process*, Vol. 1, Issue 1, pp.1-11, Jan. 2006

[4]. DIRSIG, http://dirsig.cis.rit.edu/, last visited October, 2008

[5]. Fibreoptic Systems Inc. 60 Moreland Rd, Unit A, Simi Valley, CA 93065, USA http://www.fibopsys.com/

[6]. Fletcher-Holmes, D. W. and Harvey, A. R., “Real-time imaging with a hyperspectral fovea,” *J. Opt. A: Pure Appl. Opt.* 7, 2005, pp. S298-S302

[7]. Goldberg, A.C., Stann, B., and Gupta, N., “Multispectral, hyperspectral, and three-dimensional imaging research at the U.S. Army research laboratory”, *Proceedings of the Sixth International Conference of Information Fusion*, 2003. Vol. 1, 2003, pp: 499-506

[8]. Gruninger, J., Ratkowski A. J., and Hoke, M. L., “The sequential maximum angle convex cone (SMACC) endmember model". *Proceedings SPIE, Algorithms for Multispectral and Hyper-spectral and Ultraspectral Imagery*, Vol. 5425-1, Orlando FL, April, 2004.

[9]. Harvey, A. R., and Fletcher-Holmes, D. W., *Imaging apparatus*, GB Patent application 0215248.6, 2nd July 2002 Horn, S., DARPA’s Adaptive Focal Plane Array (AFPA) Program, Available: http://www.arpa.mil/mto/programs/afpa/index.html

[10]. Headwall Photonics, Inc., 601 River Street, Fitchburg, MA 01420, USA. http://www.headwallphotonics.com/

[11]. Horn, S., DARPA’s Adaptive Focal Plane Array (AFPA) Program, Available: http://www.arpa.mil/mto/programs/afpa/index.html. Last visited 2007.

[12]. Kincade, K., “MIRTHE center aims to take mid-IR sensors to new heights. (Optical Sensing)”, *Laser Focus World*, July, 2006.

[13]. Kruse, F. A., et al., "The spectral image processing system (SIPS) - Interactive visualization and analysis of imaging spectrometer data". *Remote Sensing of the Environment*, v. 44, p. 145 - 163.

[14]. Land, M. F., and Nilsson, D.-E., *Animal Eyes*, Oxford University Press, 2002

[15]. Qu, Y., Wang, T., and Zhu, Z. Remote Audio/Video Acquisition for Human Signature Detection, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Workshop on Biometrics*, June 20-June 25, 2009, pp.66-71

[16]. Qu, Y., Wang, T. and Zhu, Z. An Active Multimodal Sensing Platform for Remote Voice Detection, *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2010)*, Montreal, July 6-9, 2010

[17]. Schott, J. R., Brown, S. D., Raqueño, R.V., Gross, H. N., and Robinson, G., "An advanced synthetic image generation model and its application to multi/hyperspectral algorithm development", *Canadian Journal of Remote Sensing*, Vol. 25, No. 2, June 1999

[18]. Schott, J. R., "Remote Sensing: the image chain approach", 2nd edition, Oxford, NY: Oxford University Press, 2007.

[19]. Soille, P., "Morphological image analysis: principle and applications", Springer Press, 1999

[20]. Start, T. M., and Fischler, M. A., "Context-based vision: recognizing objects using information from both 2D and 3D imagery", *Pattern Analysis and Machine Intelligence*, IEEE Trans. Vol. 13, Issue 10, Oct 1991

[21]. Wang, T., Zhu, Z. and Divakaran, A., Long-Rang, Audio and Audio-Visual Event Detection Using a Laser Doppler Vibrometer, *SPIE Defense, Security and Sensing: Evolutionary and Bio-Inspired Computation: Theory and Applications IV*, April, 2010

[22]. Winter, M. F., "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data." *Proceedings SPIE*, Vol. 3753, 1999, pp. 266-275

Appendix: Publications under the Support of This Grant

1. T. Wang, Z. Zhu, R. S. Krzaczek, H. E. Rhody, **A System Approach to Adaptive Multimodal Sensor Designs**, a chapter in *Machine Vision Beyond Visible Spectrum*, eds. R. Hammond, G. Fan, R. McMillan, and K. Ikeuchi, Springer, submitted in January, 2010 (accepted)
2. T. Wang, Z. Zhu, E. Blasch, **Bio-Inspired Adaptive Hyperspectral Imaging for Target Tracking,** *IEEE Sensors Journal*, Special issue on Enhancement Algorithms, Methodologies & Technology for Spectral Sensing, vol 10, no 3, March 2010, pp. 647-654
3. Z. Zhu, Y.-C. Hu and L. Zhao, **Gamma/X-Ray Linear Pushbroom Stereo for 3D Cargo Inspection**, *Machine Vision and Applications*, Volume 21, Issue 4 (2010), Page 413-425, http://dx.doi.org/10.1007/s00138-008-0173-8
4. Y. Qu, T. Wang and Z. Zhu, **Vision-aided Laser Doppler Vibrometers for Remote Automatic Voice Detection**, *IEEE/ASME Transactions on Mechatronics*, submitted on April 8, 2010.
5. Z. Zhu, **Mobile Sensors for Security and Surveillance**, *Journal of Applied Security Research*, the Haworth Press, vol 4, no 1&2:79–100, January 2009
6. H. Tang and Z. Zhu, **Content-Based 3D Mosaics for Representing Videos of Dynamic Urban Scenes,** *IEEE Transactions on Circuits and Systems for Video Technology*, accepted, August 2008.
7. Y. Qu, T. Wang and Z. Zhu, **An Active Multimodal Sensing Platform for Remote Voice Detection**, *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2010)*, Montreal, July 6-9, 2010 (accepted)
8. Y. Qu, W. Khoo, E. Molina, Z. Zhu, **Multimodal 3D Panoramic Imaging Using a Precise Rotating Platform,** *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2010)*, Montreal, July 6-9, 2010 (accepted)
9. T. Wang, Z. Zhu, and A. Divakaran, **Long-Rang, Audio and Audio-Visual Event Detection Using a Laser Doppler Vibrometer**, *SPIE Defense, Security and Sensing: Evolutionary and Bio-Inspired Computation: Theory and Applications IV*, April, 2010
10. Y. Qu, T. Wang, Z. Zhu, **Remote Audio/Video Acquisition for Human Signature Detection**, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Biometrics*, June 20-June 25, 2009, pp.66-71
11. E. Molina, Z. Zhu, O. Mendoza-Schrock, **Mosaic-based 3D scene representation and rendering of circular aerial video**, *SPIE Defense, Security and Sensing: Evolutionary and Bio-Inspired Computation: Theory and Applications IV*, April, 2010
12. T. Wang, Z. Zhu, H. Rhody, **A Smart Sensor with Hyperspectral/Range Fovea and Panoramic Peripheral View**, pp.98-105, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Object Tracking and Classification Beyond and in the Visible Spectrum*, June 20-June 25, 2009

13. H. Tang, Z. Zhu and J. Xiao, **Stereovision-Based 3D Planar Surface Estimation for Wall-Climbing Robots**, *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 11-15, 2009, St. Louis, USA

14. T. Wang and Z. Zhu, **Intelligent Multimodal and Hyperspectral Sensing for Real-Time Moving Target Tracking**, *37th IEEE Applied Imagery Pattern Recognition Workshop (AIPR'08): Multiple Image Information Extraction*, Cosmos Club, Washington DC, October 15-17, 2008